YUN-YUN (ALICE) TSAI

(+1) 917-498-2709 • ✉ yt2781@columbia.edu • https://yunyuntsai.github.io

RESEARCH INTERESTS

My research focuses on **machine learning robustness in vision + language domain**, domain adaptation, self-supervised learning, and **GenAI safety**. My personal goal is to build trustworthy and reliable ML systems.

EDUCATION

Columbia University in the City of New York — New York, USA
Ph.D. student in Computer Science — *Sept., 2021 – Present*
– Advisor: Prof. Junfeng Yang

National Tsing Hua University — Hsinchu, Taiwan
Master of Science in Computer Science — *Sept., 2018 – June, 2020*
– Advisor: Prof. Tsung-Yi Ho
Bachelor of Science in Computer Science — *Sept., 2014 – June, 2018*

RESEARCH AND WORKING EXPERIENCE

Research Scientist Intern — **Meta, Superintelligence Labs**, Bellevue, WA
Mentor: Sinong Wang , Yuning Mao — *May, 2025 – present*
- Project: Benchmarking LLMs for WebUI Code Generation with UI Agent Evaluation

AI Research Scientist Intern — **Meta, GenAI Team**, New York, NY
Mentor: Laurens van der Maaten , Chuan Guo — *May, 2024 – Aug, 2024*
- Project: RAFP: Identifying LLM Lineages via Rare-Region Fingerprints.

Applied Scientist Intern — **Amazon, CoRo Team**, Bellevue, WA
Mentor: Che-chun Su, Fu-chen Chen, Min Sun — *June, 2023 – present*
- Project: Improving vision robustness on out-of-domain (OOD) data via stable diffusion.

PUBLICATIONS

Conference and Workshop Papers

[C1] **Yun-Yun Tsai**, Yuning Mao, Shiqi Wang, Junfeng Yang, Sinong Wang, " WebUIProof: Benchmarking WebUI Code Generators with UI-Agent Execution Harness ," (**arXiv** Preprint) 2026.

[C2] Chunyi Wang, Yunfei Ke, Junfeng Yang, **Yun-Yun Tsai**, Penghui Li, " ARQ: Agentic CodeQL Query Refinement for C/C++ Vulnerability Detection," (**arXiv** Preprint) 2026.

[C3] Jincen Song, Luke Tenyi Wang, **Yun-Yun Tsai**, Penghui Li, and Junfeng Yang, " Reasoning under Vision: Understanding Visual-Spatial Cognition in Vision-Language Models for CAPTCHA," in Findings of the Association for Computational Linguistics: (**AACL**) 2026.

[C4] **Yun-Yun Tsai**, Jia-Hao Liang, Chuan Guo, Junfeng Yang, Laurens van der Maaten, "RAFP: Identifying LLM Lineages via Rare-Region Fingerprints," (**arXiv** Preprint) 2025.

[C5] **Yun-Yun Tsai**, Qingyuan Liu, Ruijian Zha, Victoria Li, Pengyuan Shi, Chengzhi Mao, Junfeng Yang, " REVEAL: Robust Evolution of Vision-Language Models for Explainable AI-Video Detection," (**ICCV**) 2025, Knowledge-Intensive Multimodal Reasoning Workshop

[C6] **Yun-Yun Tsai**, Fu-Chen Chen, Albert Y.C. Chen, Junfeng Yang, Che-Chun Su, Min Sun, Cheng-Hao Guo, "GDA: Generalized Diffusion for Robust Test-time Adaptation," in Proceeding of the IEEE / CVF Computer Vision and Pattern Recognition Conference (**CVPR**) 2024.

[C7] **Yun-Yun Tsai**, Ruize Xu, Chengzhi Mao, Junfeng Yang, "From Detection to Deception: Are AI-Generated Image Detectors Adversarially Robust?," **CVPR** 2024, Responsible Generative AI Workshop.

[C8] Qingyuan Liu, Pengyuan Shi, **Yun-Yun Tsai**, Chengzhi Mao, Junfeng Yang, "Turns Out I'm Not Real: Towards Robust Detection of AI-Generated Videos," **CVPR** 2024, Generative Models for Computer Vision workshop.

[C9] **Yun-Yun Tsai**, Chengzhi Mao, Junfeng Yang, "Convolutional Visual Prompt for Robust Visual Perception," 37th Conference on Neural Information Processing Systems (**NeurIPS**) 2023.

[C10] **Yun-Yun Tsai**, Ju-Chin Chao, Albert Wen, Zhaoyuan Yang, Chengzhi Mao, Tapan Shah, Junfeng Yang, "Test-time Detection and Repair of Adversarial Samples via Masked Autoencoder," in Proceeding of the IEEE / CVF Computer Vision and Pattern Recognition Conference (**CVPR**) 2023, AdvML workshop

[C11] Lei Hsiung, **Yun-Yun Tsai**, Pin-Yu Chen, and Tsung-Yi Ho, "Towards Compositional Adversarial Robustness: Generalizing Adversarial Training to Composite Semantic Perturbations," in Proceeding of the IEEE / CVF Computer Vision and Pattern Recognition Conference (**CVPR**) 2023.

[C12] **Yun-Yun Tsai**, Lei Hsiung, Pin-Yu Chen, and Tsung-Yi Ho, "Generalizing Adversarial Training to Composite Semantic Perturbations," in Proceeding of International Conference on Machine Learning (**ICML**) AdvML Workshop, 2021.

[C13] Lei Hsiung, **Yun-Yun Tsai**, Pin-Yu Chen, and Tsung-Yi Ho, "CARBEN: Composite Adversarial Robustness Benchmark," in Proceeding of International Joint Conference on Artificial Intelligence (**IJCAI**), 2022.

[C14] Chao-Han Huck Yang, **Yun-Yun Tsai**, and Pin-Yu Chen "Voice2Series: Reprogramming Acoustic Models for Time Series Classification," in Proceeding of International Conference on Machine Learning (**ICML**), 2021.

[C15] **Yun-Yun Tsai**, Pin-Yu Chen, and Tsung-Yi Ho, "Transfer Learning without Knowing, Reprogramming black box machine learning model with scarce data and limited resources," in Proceeding of International Conference on Machine Learning (**ICML**), 2020.

[C16] **Yun-Yun Tsai**, Pin-Yu Chen, Tsung-Yi Ho, "Adversarial Machine Learning for Social Good: Reprogramming black box machine learning model with scarce data and limited resources," Advances in Neural Information Processing Systems (**NeurIPS**) NewInML Workshop, Poster, 2019.

[C17] Honggang Yu, Kaichen Yang, Teng Zhang, **Yun-Yun Tsai**, Tsung-Yi Ho, Yier Jin, "CloudLeak: Large-Scale Models Extraction Through Adversarial Examples," in Proceeding of Network and Distributed System Security Symposium (**NDSS**), 2020.

[C18] Ta-Wei Huang, **Yun-Yun Tsai**, Chung-Wei Lin, Tsung-Yi Ho, "Vehicle Sequence Reordering with Cooperative Adaptive Cruise Control," in Proceeding of Design, Automation and Test in Europe Conference and Exhibition (**DATE**), 2019.

Patents

[P1] Pin-Yu Chen, **Yun-Yun Tsai**, Sijia Liu, Chia-Yu Chen, I-Hsin Chung, Tsung-Yi Ho. "Transfer Learning With Machine Learning Systems", U.S. Patent Application No: 17/029506, Application Date: September 23, 2020.

[P2] Pin-Yu Chen, I-Hsin Chung, Bo Wu, Chuang Gan, Lei Hsiung, **Yun-Yun Tsai**, Tsung-Yi Ho. "Composite Adversarial Attack Model Training For Neural Networks", U.S. Patent Application No: 18/331,211, Application Date: May 12, 2026.

Honors, Awards, and Grants

Student Travel Grant, CVPR 2024, Seattle, Washington.

Scholar Award, NeurIPS 2023, New Orleans, Louisiana.

Student Travel Grant, CVPR 2023, Vancouver, Canada.

Ph.D. Dean's Fellowship, Fu Foundation School of Engineering and Applied Sciences, Columbia University, 2021.

Best Presenter, Blackhat Award Forum in CyberSec, Taiwan, 2020.